CPS^2: a Contextual Privacy Framework for Social Software

Rula Sayaf[1], Dave Clarke[1,2], and Richard Harper[3]

[1] iMinds-DistriNet, Dept. of Computer Sciences, KU Leuven, Belgium
Email: rula.sayaf@cs.kuleuven.be
[2] Dept. of Computer Sciences, Uppsala University, Sweden
Email: dave.clarke@cs.kuleuven.be, dave.clarke@it.uu.se
[3] Microsoft Research, Cambridge, United Kingdom
Email:r.harper@microsoft.com

Summary. Social software has become one of the most prominent means for communication. Context is essential for managing privacy and guiding communication. However, in social software, context can be ambiguous due to the overload of data and the mix of various audiences, resulting in privacy issues.
To overcome context issues, we analyse the role of context in communication and privacy management, and propose CPS^2, a conceptual framework of contextual privacy management. CPS^2 identifies the interpretation of data as the ingredient in contextual privacy management that once preserved within any disclosure context, contextual privacy can be preserved. We present how CPS^2 can be technically realised, and how it addresses context issues and offers fine-granular context control.

1 Introduction

Communication through social software is becoming one of the most prominent ways of daily communication. Social software is application software for the exchange of various types of data and communication with a large number of users. Such communication is simple as it can be achieved by disclosing data to other users. This simplicity comes often at a price in terms of privacy issues. Issues specifically occur when data is accessed by inappropriate audience or put in in inappropriate contexts [1]. To mitigate such issues, users should have means to preserve privacy in a context-based manner to ensure the appropriateness of disclosure contexts.

Context is essential for both communication and privacy management [2, 3]. Through communication, the interlocutors express their identity with the data they disclose. Besides its other dimensions, privacy concerns controlling data flow to manage one's identity [4, 5]. When context is unclear communication can be disrupted affecting one's identity expression and privacy. In social software, communication is characterised by ambiguous context due to the mixing of different audiences from different contexts. As a result, managing privacy and communicating based on context can be challenging.

Managing privacy through controlling context is a complex task. Controlling context requires reasoning about the current context and how it may change [6].

Such reasoning is challenging due to the high-dimensionality of context parameters [7]. Current context-based privacy management approaches address such complexity by simplifying context representation resulting in a limited control over context [8]. To understand the insufficiency of context-based management consider the following scenario that is based on a reported incident of 'Antwerpen hoeren' (prostitutes of Antwerpen) [9]:

Scenario 1 Els is a fashion model, and she posts her photo in a swimming suit on Facebook and makes it public. Although it is viewed by public, Els experiences a privacy issue when her photo is posted in the context of 'prostitutes of Antwerpen' page, which affects her job applications. In contrast, Els does not face any issue when her photo is put in 'jobs for top models ' page.

Current privacy management approaches do not offer context control to mitigate such violations. Rather, for example, they offer control over the type of audience who can access and handle data [8]. In this paper, we address context control issues by proposing a conceptual framework for contextual privacy management. We analyse the context-privacy relation and argue that the interpretation of data is the ingredient that captures this relation, and that by ensuring the integrity of interpretation, contextual privacy can be managed. The framework is a conceptual approach to manage privacy in different contexts without burdening users with reasoning about context and its complexities. The contributions of this paper are manifold:

1. Analysing the problems of controlling data and managing privacy in a context-based manner (Section 2)
2. Analysing the role of context in privacy and communication (Section 3)
3. Proposing a conceptual framework for Contextual Privacy for Social Software (CPS^2), and presenting how this framework can be technically realised and can address context and privacy issues (Section 4).

2 Problem Statement

Communicating while preserving privacy in any data disclosure context requires a fine-grained control of context [8]. In social software, context identifies situations where various types of data are disclosed and users interact. Context ambiguity is one of the main issues in social software communication. Ambiguity means that it is challenging to accurately identify the current context. Ambiguity obstructs the clarification of the communicative message, and the user's assessment of her privacy.

Privacy management can be challenging due to context management problems. Privacy is viewed as the means to control the disclosure of data within boundaries and controlling data disclosure contexts [10, 11]. By following these views, contextual privacy management requires two types of control: control over the *original* context in which the data was originally disclosed through the software, and control over any disclosure context by specifying appropriate or inappropriate contexts. Facilitating these two types of control is rather challenging. A user can control the original context by choosing where to disclose

data and to whom. However, over time, the original context might change and evolve [6] into an inappropriate context. In order to avoid such situations, users should constantly monitor context change. Such monitoring is challenging because users do not invest much time in managing online communication [12], and it is especially challenging when context is ambiguous. Having control over any possible disclosure context requires listing possible appropriate or/and inappropriate disclosure contexts, depending on the assumed closed- or open-world of contexts. Given the 'theoretically infinite complexity' of social situations, and the infinite set of possible contexts [13, 14], it may be infeasible to list of all possible contexts [15]. These issues are often mitigated by simplistic context representation in privacy management approaches that offer coarse-grained context control [8]. For instance, contexts can be captured by the roles of users to make it easier to list possible appropriate or inappropriate contexts in a system.

3 Analysis of Context and Privacy

Context is the information construct that characterises the communication situation [7]. Context is a container of data; it enables inferring the relevant meaning of the communicative message [7]. A data item can have a set of different possible meanings or interpretations, and by identifying the context it is put within, the relevant interpretation can be inferred. For example, the page in which Els's photo is put is a context related to 'prostitutes', this context is inferred by information about the social software type, page type, page content, creator of the page, page name and other meta data. When Els's photo is put in this context, the most relevant meaning is a 'prostitute_photo'.

In online communication, privacy management can be a mean for identity management [5]. The data owner[1] discloses a data item to communicate about it with the selected audience. Through communication, the owner expresses a specific identity and manages it by specifying who the audience are and what data they could access in a specific context [16]. The owner needs to be aware of how others would perceive and interpret a data item to make the privacy decision of to whom disclose it [5]. Thus, the interpretation of the data and context are of central roles in the privacy management process.

The importance of context and the interpretation of data can be observed in communication types: cooperative and adversarial. Such types are the extreme ends of the communication spectrum that are characterised by variant degrees of trust, context involvement, and privacy concerns [17]. In *cooperative communication*, the interlocutors trust each other [18] and act jointly to understand the communicated message. Cooperative communication can be achieved by following the Gricean maxims, which are providing a sufficient amount of information that are true, relevant, and unambiguous to make context explicit [19]. Abiding by those maxims can be challenging in ambiguous contexts. In an *adversarial communication*, an interlocutor—the adversary—acts maliciously and misleads others into misinterpreting the message to disrupt communication. Context am-

[1] We do not imply the legal ownership.

biguity hinders the correct interpretation of data and may result in unintended adversarial communication where privacy concerns are high.

Based on the above-mentioned argument, we define context-based or contextual privacy management as the process of making disclosure decisions that maintain the appropriate interpretation of the owner's data, in order to express the owner's desired identity in a specific context. To achieve that, context clarity is essential. However, clarity of context requires effort to make communication cooperative and avoid adversarial communication. To facilitate contextual privacy management and avoid overloading user with context complexities, we propose CPS^2 in the following section.

4 CPS^2: Contextual Privacy for Social Software

The main idea of CPS^2 is to facilitate communication with an increased level of privacy without burdening users with context management.

We propose CPS^2 to manage contextual privacy by managing the interpretation of data. Rather than simplifying the representation of context or imposing reasoning about context on users to specify privacy management policies, and given the technological advances in context inference [20] and automatic data interpretation [21], the framework proposes lifting the burden of reasoning about context to the level of the social software platform, and allowing owners to state the appropriate interpretation of their data. Accordingly, the framework guards the appropriate interpretation upon any change of context.

To understand CPS^2 consider scenario 1: Els's profession as a fashion model is indicated on her page, thus, the context of her profile page indicates that the 'fashion-related' interpretation is the most relevant interpretation. Upon viewing the photo, the audience would highly likely perceive the interpretation of the photo as such, although there is no guarantee what the interpretation the audience would subjectively infer. When the photo is put in the 'prostitutes' context, the relevance of the 'fashion-related' interpretation is low and the relevance of the 'prostitute' interpretation is high, which affects El's identity. With CPS^2, Els can specify the set of appropriate interpretations of the photo as {fashion_show, swim_suits_show, pretty_model}. Accordingly, the recontextualisation into the 'prostitutes' context is prohibited because it results in an interpretation that is not in the set Els has specified, while the recontextualisaiton into the 'jobs for top models' context is allowed.

4.1 Realisation of CPS^2

The realisation of CPS^2 implies a system with three main functions: context inference, interpretation inference, and contextual management. CPS^2 assumes the existence of an underlying context inference and interpretation inference layers that need not be managed by users, but by the social software provider, for instance. The realisation would comprise the following layers:

1. Context inference layer: responsible for inferring or labelling the context of the current situation within the social software realm. The input to this layer

is the social software data: users and their attributes, data items, relations, ads, and the structure of its pages and modules. When data is added to a situation, this layer adapts and infers the new context.
2. Interpretation inference layer: responsible for inferring the interpretation of data based on the context inferred by the previous layer. The data can be interpreted whether it is textual or visual.
3. CPS^2 control layer: responsible for facilitating contextual privacy management by means of two possible approaches: access control or accountability and auditing approach. The access control approach comprises a policy language to express the contextual privacy policies and an enforcement mechanism. A policy can be formulated to express the appropriate interpretations of a data item . Upon performing an action—resulting in adding or removing data from a context—the control layer consults the policies of data items in that context and verifies the interpretation inferred by the previous layer. The action is performed only if no policies are violated.
   In the accountability and auditing approach, users need not specify policies. Rather, upon each context change, the framework marks the actions that cause a change of the original interpretation in the original disclosure context. The data owner then could verify the appropriateness of the new interpretation.

4.2 Addressing Context and Privacy Problems

CPS^2 could potentially address the problems mentioned in Section 2, as follows:

1. Context ambiguity: the framework addersses this problem not by making the context less ambiguous to users, but rather, even if context is ambiguous to users, only appropriate actions are allowed because the context inference layer could still identify context given all the data it has more accurately than users can.
2. Context simplistic representation: by shifting the burden of reasoning about context to the underlying framework. This guarantees that the context richness can indirectly be employed to manage users' privacy.
3. Control over the original context: by facilitating the management of interpretation, owners can indirectly control context to a relatively high degree without having to monitor the changes of context and the possible violations.
4. Control over any disclosure context: the previous argument is valid here. The framework facilitates effortless control over any context by continuously monitoring the interpretation in any context data is put in.

Moreover, CPS^2 enhances communication to become cooperative even if context is ambiguous, by allowing only appropriate actions that may not affect the interpretation of data. It also facilitates avoiding adversarial communication by preserving data interpretation. CPS^2 facilitates control over data flow in private or public spaces and allows disclosures that preserve users' identities.

5 Related Work

Many works have incorporated context in privacy management. On the conceptual level, Nissenbaum proposes contextual integrity [22] for privacy management. She presents a list of norms: contexts, actors, attributes, and transmission principles, that must be maintained to preserve privacy. Our framework differs from this theory by not requiring an exhaustive specification of the possible contexts or the other ingredients in the theory. The complexity of contextual integrity results in models that adopt simplistic context representation to overcome the complexity, such as the formal model of Barth *et. al.* [23] where context are represented by roles of users. Similarly, Fong proposed a social software-specific access control model in which relationships are viewed as contexts [24]. In contrast to CPS^2, Fong's model offer control over the original context but not over any data disclosure context. Moreover, the simplification of these models reduces the granularity that contexts offer and fails in addressing the problems discussed in Section 2.

6 Conclusion and Future Work

In CPS^2, we propose maintaining data interpretation to manage contextual privacy and address the complexity of controlling context. By this utilisation, the richness of context can be indirectly employed in managing privacy, while users can specify simple policies about the appropriate interpretation. CPS^2 enhances communication in which interpretation is essential. In other work, we have conducted some experiments related to context inference, and we will report them elsewhere. Our future work aims at providing a realisation of the framework.

Acknowledgment

This research has been funded by the IWT in the context of the SBO project on Security and Privacy for Online Social Networks (SPION). Thanks are due to Natasa Milic-Frayling and Sören Preibusch at Microsoft Research Cambridge.

References

1. Goldie, J.: Virtual communities and the social dimension of privacy. University of Ottawa Law & Technology Journal **3**(1) (2003) 133–167
2. Clark, H., Carlson, T.: Context for comprehension. In: J. Long & A. Baddeley (Eds.). Lawrence Erlbaum Associates, Inc. Hillsdale, NJ (1981) 313–330
3. Majeski, M., Johnson, M., Bellovin, S.M.: The Failure of Online Social Network Privacy Settings. Technical Report CUCS-010-11, CS, Columbia University (2011)
4. Gürses, S.: Multilateral privacy requirements analysis in online social network services. PhD thesis (2010)
5. Palen, L., Dourish, P.: Unpacking "privacy" for a networked world. In: Proceedings of the SIGCHI conference on Human factors in computing systems. CHI '03, New York, NY, USA, ACM (2003) 129–136

6. Mcculloh, I.: Detecting changes in a dynamic social network (social networks, change detection, network dynamics). PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA (2009)
7. Van Dijk, T.A.: Discourse and context. A Sociocognitive Approach, Cambridge University (2008)
8. Sayaf, R., Clarke, D.: Access control models for online social networks. In Caviglione, L., Coccoli, M., Merlo, A., eds.: Social Network Engineering for Secure Web Data and Services, IGI (2013) 32–65
9. De Wolf, R.: Over 'spotted', 'hoeren' en 'failed'-pagina's. Electronic article: http://www.knack.be/nieuws/belgie/dader-antwerpse-hoeren-fotogeklist/article-4000230766578.htm, Last checked Feb. 2013 (2013)
10. Westin, A., Blom-Cooper, L.: Privacy and freedom. Atheneum New York (1970)
11. Petronio, S.: Boundaries of privacy: Dialectics of disclosure. SUNY Press (2002)
12. Lipford, H.R., Besmer, A., Watson, J.: Understanding privacy settings in facebook with an audience view. In: Proceedings of the 1st Conference on Usability, Psychology, and Security, Berkeley, CA, USA, USENIX Association (2008) 2:1–2:8
13. Van Dijk, T.A.: Context models in discourse processing. The construction of mental representations during reading (1999) 123–148
14. Skantze, G.: Error Handling in Spoken Dialogue Systems-Managing Uncertainty, Grounding and Miscommunication. Doctoral dissertation, KTH. PhD thesis, Department of Speech, Music and Hearing (2007)
15. Lampinen, A., Lehtinen, V., Lehmuskallio, A., Tamminen, S.: We're in it together: interpersonal management of disclosure in social network services. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM (2011) 3217–3226
16. Wood, A.F., Smith, M.J.: Online communication: Linking technology, identity, & culture. Routledge (2004)
17. Harper, R.H.: Texture: human expression in the age of communications overload. MIT Press (2010)
18. Harper, R., ed.: Trust, Computing and Society. CUP: New York (2014)
19. Grice, H.P.: Logic and conversation. In Davidson, D., Harman, G., eds.: The Logic of Grammar. Harvard Univ. (1975) 64–75
20. Cao, H., Hu, D.H., Shen, D., Jiang, D., Sun, J.T., Chen, E., Yang, Q.: Context-aware query classification. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, ACM (2009) 3–10
21. Celikyilmaz, A., Hakkani-Tur, D., Tur, G.: Statistical semantic interpretation modeling for spoken language understanding with enriched semantic features. In: Spoken Language Technology Workshop (SLT), 2012 IEEE, IEEE (2012) 216–221
22. Nissenbaum, H.: Privacy in context: Technology, policy, and the integrity of social life. Stanford Law & Politics (2010)
23. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: Framework and applications. In: Proceedings of the 2006 IEEE Security & Privacy. Number 15 in SP '06, IEEE, IEEE Computer Society (2006) 184–198
24. Fong, P.W.L.: Relationship-based access control: protection model and policy language. In: Proceedings of the first ACM conference on Data and application security and privacy. CODASPY 11, New York, NY, USA, ACM (2011) 191–202